iCAUSALBAYES USER MANUAL

INTRODUCTION

You can use this app to build a causal Bayesian network and experiment with inferences. We hope you'll find it interesting and helpful.

We expect most of our users will be AI students with a basic knowledge of probability and statistics. You don't need a deep understanding of regression to use the app but you do need to be aware of some terminology and the basic regression equation: a *response variable* (RV) *depends* on one or more *explanatory variables* (EV) plus an *error term* (ET).

The basic regression equation, then, is:

$$RV = C_1EV_1 + C_2EV_2 + \cdots + C_nEV_n + ET$$

CREATING A NETWORK: A TUTORIAL

There are two ways to create a network:

1) Build a network in the app. (recommended)

2) Create and import an xml or zip file describing the network.

IMPORTING A NETWORK

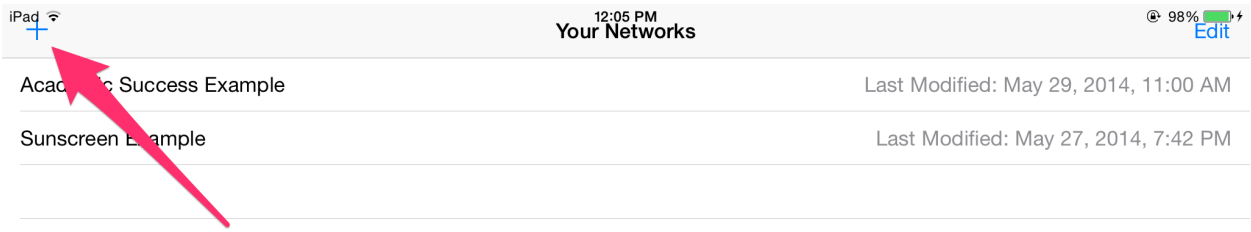
This app is capable of importing a network by opening a properly formatted zip or xml file from another app. By pressing the Email button, the app can generate and send zip files of the required format. After receiving the email, select the attachment and choose "Open in" then select iCausalBayes.

It is possible to create your own network by writing or editing an xml file with a text editor then zipping that file along with a folder containing any images the network should use (if any are used). To see the format of these files, use the app to email any existing network to yourself, then unzip the attachment. We recommend using the app to create networks, as writing your own xml document can be very error prone.

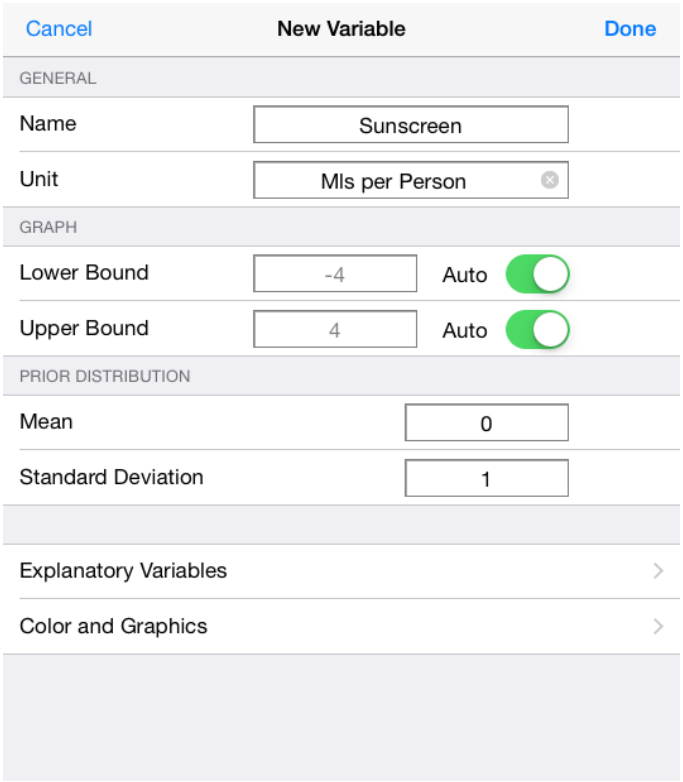
BUILDING A NETWORK IN THE APP

We'll use the same 3-node throughout this manual. This network comes bundled with the app, but, here, we'll build it from scratch.

Tap the + button at the top right of the home ('Your Networks') screen to create a new network.



Touch and hold anywhere on the screen to create a new variable at that location. Use the New Variable view to configure the variable.



Give your variable a name: `Sunscreen`.

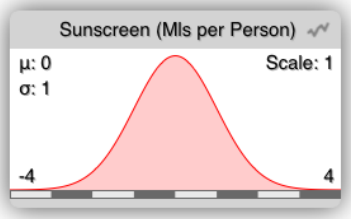
For the Unit, enter `Mls per Person`.

You can enter a Lower and Upper Bound for the graph or leave these (one or both) on Auto, This determines how many Standard Deviations will be displayed for the variable. We'll leave these on Auto. You can always come back and change this. In fact, you can always come back and change anything!

For the Prior Distribution, leave the Mean and Standard Deviation at `0` and `1`, respectively.

Tap Done in the top right of the New Variable view to apply these changes.

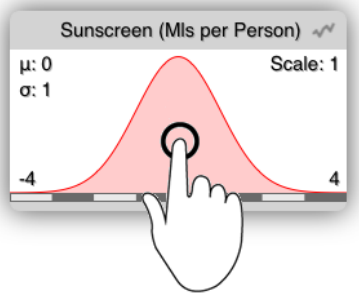
You should see the `Sunscreen` variable on your screen. The Sunscreen node shows that Mls per Person is approximately normally distributed, with a mean of 0. At this point, this node says, "on average, people use *x* mls of sunscreen".



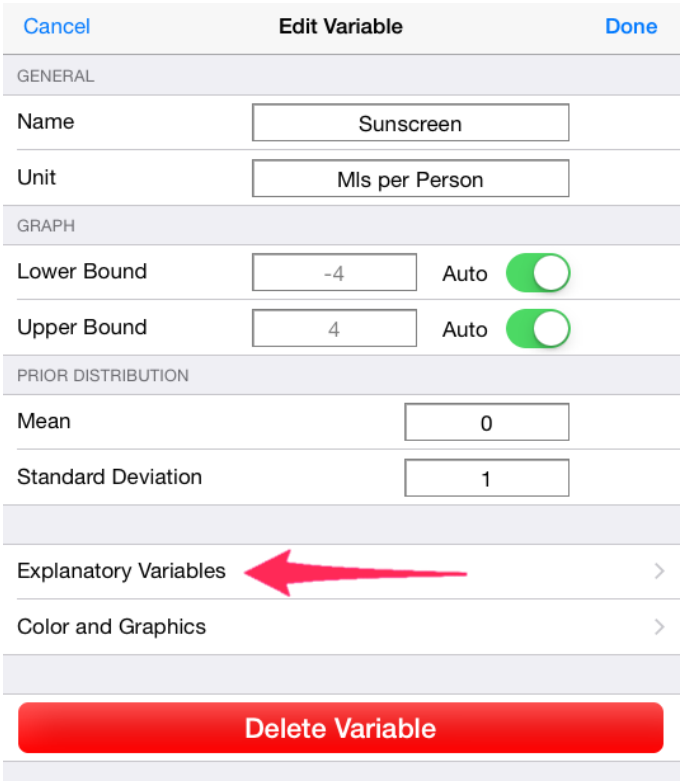
Now let's add another variable. This one is will be an Explanatory Variable for `Sunscreen`. We'll call it `Melanin Index`. This variable is a measure of skin pigmentation, as determined by a reflectometer. For the Unit, we'll use `Index`. Again, we'll leave the Upper and Lower Bounds on Auto. Again, we'll use the defaults for the Prior Distribution.

Position the variables as you like. Use the familiar 'two finger pinch' inside a node to make it bigger or smaller or outside to expand or contract the network as a whole.

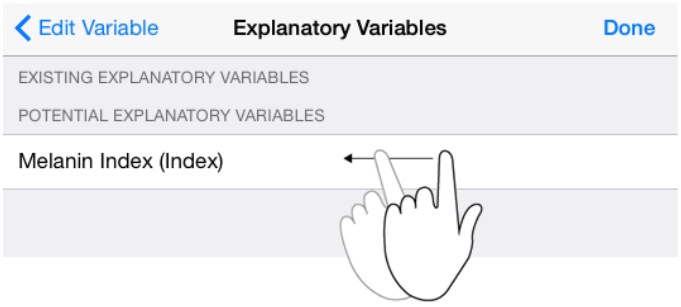
Now, we'll make `Melanin Index` an explanatory variable for `Sunscreen`. Touch and hold inside `Sunscreen` to edit the variable.



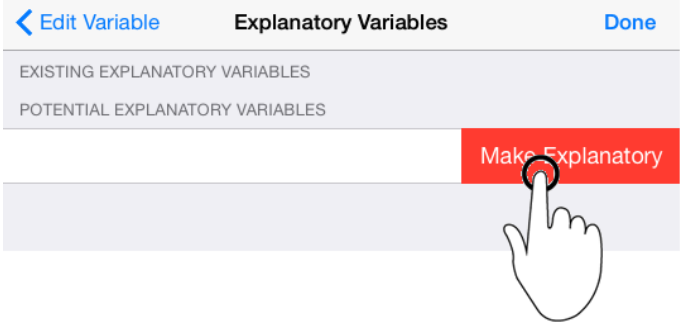
In the Edit Variable view, tap Explanatory Variables.



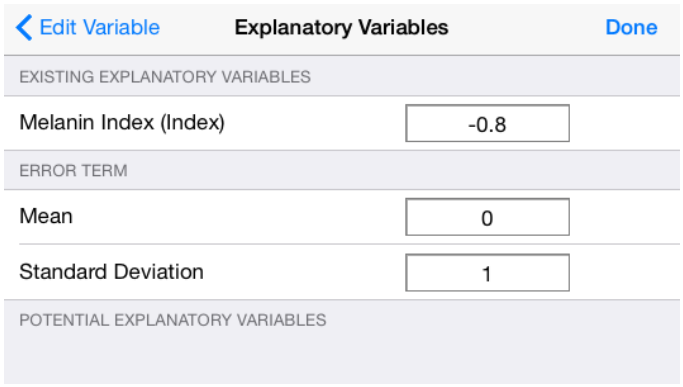
In the Explanatory Variables view, swipe left to expose the Make Explanatory button.



Tap the button to make `Melanin Index` an Explanatory Variable.

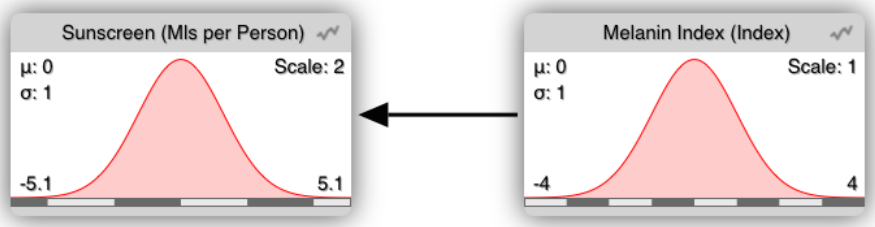


Melanin Index will move from the Potential Explanatory Variables to Existing Explanatory Variables and an Error Term will appear.



Enter `-0.8` as the coefficient for `Melanin Index`. Leave the Error Term Mean and Standard Deviation at their default values. Tap Edit Variable in the upper left corner to return to the previous screen.

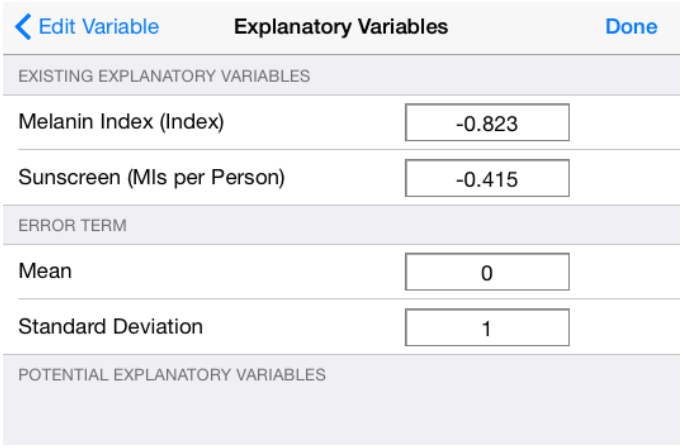
Back in the Edit Variable view, you will notice that the Prior Distribution is gone. Tap Done.



Now there is an arrow from `Melanin Index` to `Sunscreen`. We have changed the meaning of both variables. The arrow indicates that `Melanin Index` affects `Sunscreen` use. The coefficient of `Melanin Index` tells us how: people with a lower `Melanin Index` use more `Sunscreen`. This is something a Bayes Net learning algorithm might discover.

Our third and last variable will be called `Melanoma`. For Units, we'll use `Incidence per 100`. Once again, we'll leave Upper and Lower Bounds on Auto.

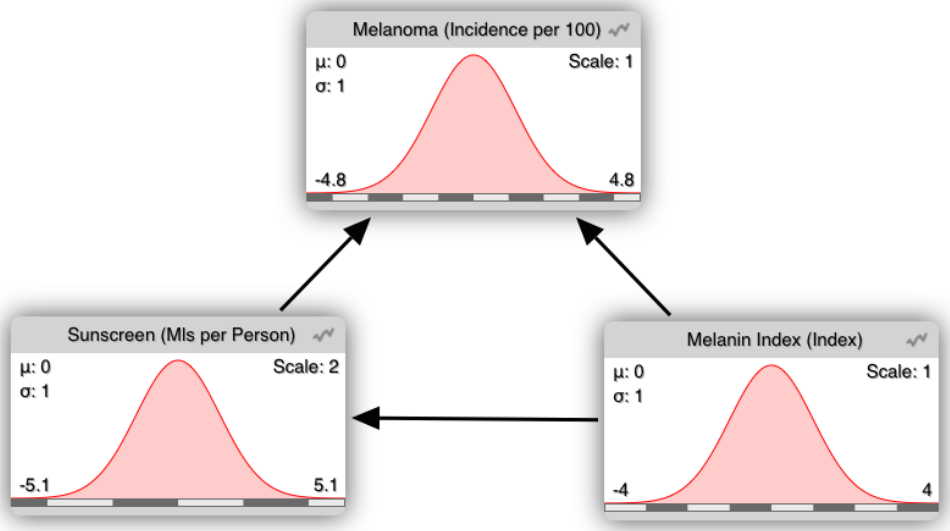
Since we have already created the Explanatory Variables for `Melanoma`, we can ignore the Prior Distribution and set these up right away. Tap Explanatory Variables, swipe to make both `Sunscreen` and `Melanin Index` explanatory variables, and enter `-.823` as the coefficient for `Melanin Index` and `-.415` as the coefficient for `Sunscreen`. We'll use the Error Term Mean and Standard Deviation defaults here, too. Tap Done in the top right corner to close the menu and add the variable to the network.



Now we have arrows from Sunscreen and Melanin Index to Melanoma. This is something that might be learned by an ordinary regression analysis, and it says that the incidence of melanoma varies with both `Melanin Index` and `Sunscreen` use.

Arrange and resize the variables so you can see everything. Save the network if you like by tapping Save As and entering a suitable name. At this point, you may also wish to set the Display Options. (You can set Display Options for the network as a whole or for each node individually.)

Our small network is complete.



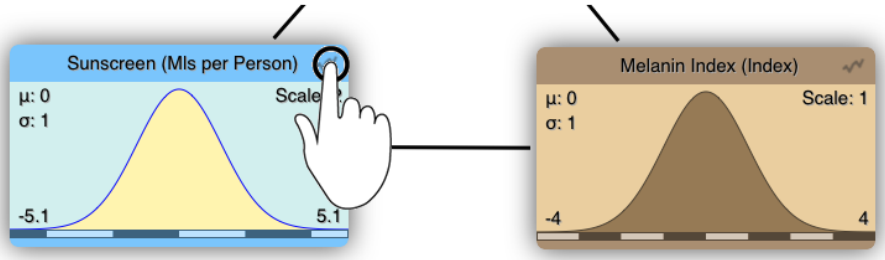
OBSERVING AND DOING

There is a long-standing controversy over sunscreen. There are those who say that people who use sunscreen end up spending more time in the sun and that the increased sun exposure negates - and even outweighs - any protection sunscreen might provide. Some argue that sunscreen users have lower levels of vitamin D, which protects against skin cancer (as well as other cancers). Others maintain that some sunscreen ingredients increase the risk of melanoma (and other cancers). On the other hand, most dermatologists and cancer organizations say that if you want to reduce your risk of melanoma, you should use sunscreen consistently.

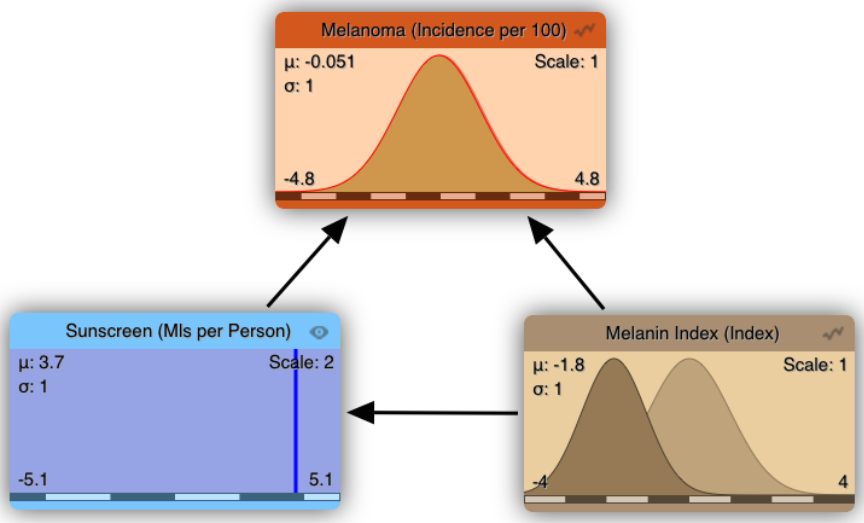
Suppose the observational data suggests that sunscreen use doesn't much affect the probability that you will develop melanoma. Now let's investigate.

OBSERVING

Tap the icon on the upper right corner of the `Sunscreen` variable to put it into Observing mode - you know you are in Observing mode when you see an eye icon in the upper right of the variable.

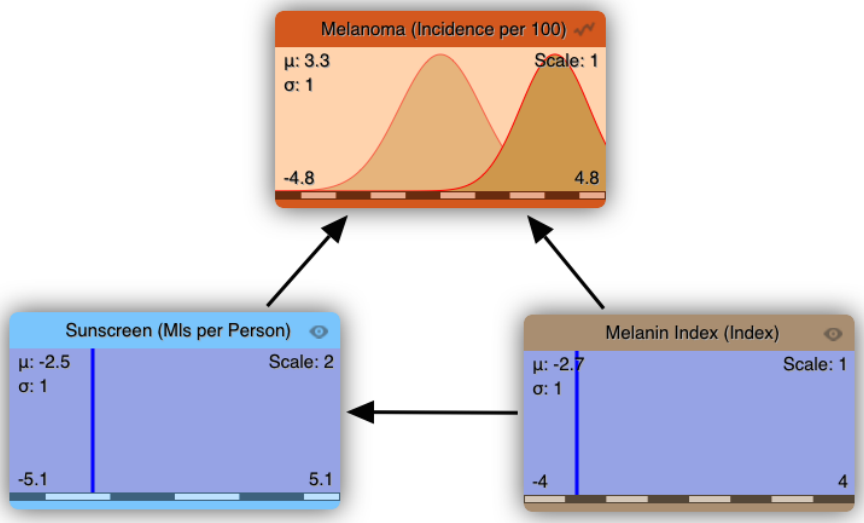


A blue line, representing an observed value, will appear in place of the distribution curve. Leave the other two variables in Graph mode (indicated by a sparkline). Now drag the line around to see the values of the other two variables for a given value of the `Sunscreen` variable. You will see that the distribution for `Melanoma` doesn't move very much. It would seem that `Sunscreen` use doesn't much affect the incidence of `Melanoma`.

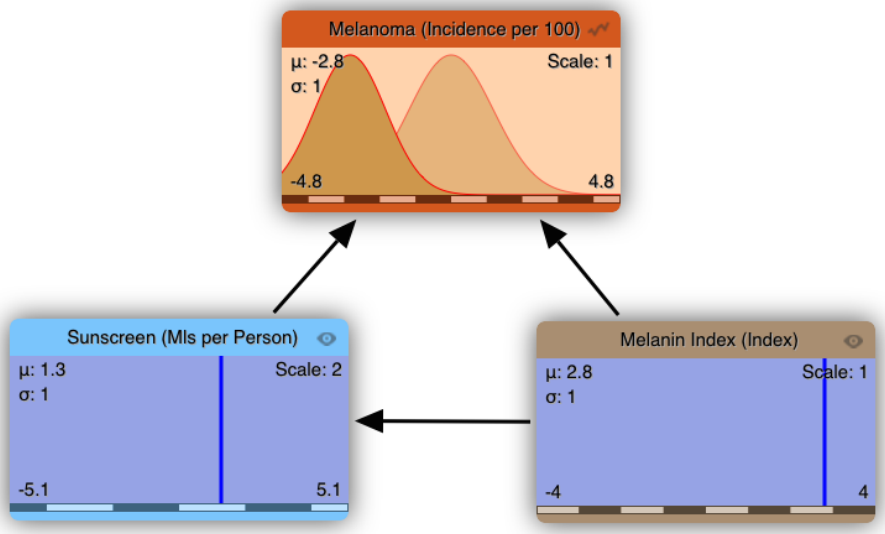


However, you might also observe something we mentioned earlier when we set up `Melanin Index` as an explanatory variable for `Sunscreen`: when `Sunscreen` goes up, `Melanin Index` goes down; when `Sunscreen` goes down, `Melanin Index` goes up. It appears that people with darker skin use less `Sunscreen` overall, while people with lighter skin use more; intuitively, this makes sense.

So what happens if we observe the effect of different values for `Melanin Index`? Tap to Observe the `Melanin Index` variable and drag the observed value line to the left. We are observing values for people with little skin pigmentation. Now drag the line for `Sunscreen` to the right. It appears that, for people with little skin pigmentation, as `Sunscreen` usage increases, the incidence of `Melanoma` goes down. Conversely, as `Sunscreen` usage decreases, the incidence of `Melanoma` goes up.



Let's see what happens when `Melanin Index` has a higher value. Drag the line to the right and then observe the effect of different values for `Sunscreen` on `Melanoma`. Again, more `Sunscreen` means less `Melanoma` and less `Sunscreen` means more `Melanoma`, though the mean has moved to the left.



It would seem that `Melanin Index` is a confound and that `Sunscreen` use reduces `Melanoma` rates.

It will be useful to consider how the probability of `Melanoma` (M), given a particular observed value of `Sunscreen` (S), is calculated. For the sake of simplicity, we'll pretend we're dealing with discrete, binary variables. (D and ¬D represent dark skin and light skin, respectively.) The calculation is as follows:

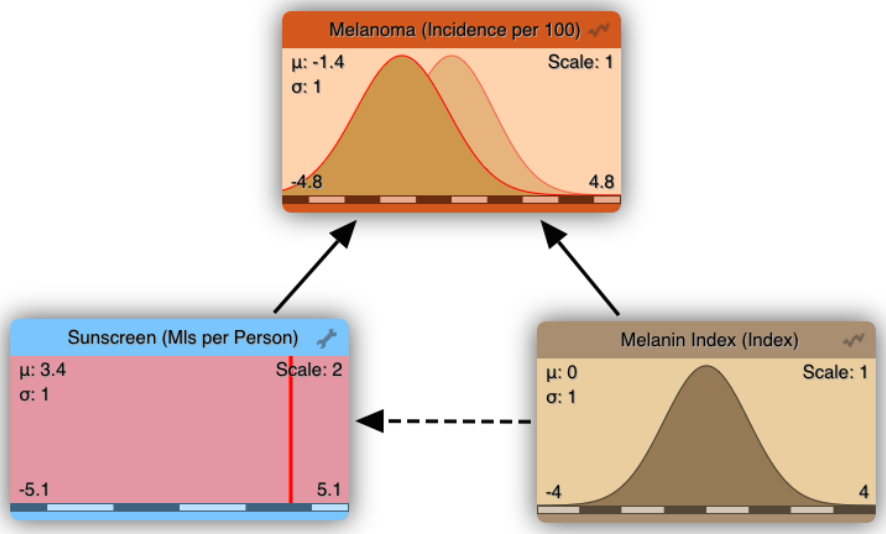
$$P(M|S) = P(M|DS)P(D|S) + P(M|\neg DS)P(\neg D|S)$$

DOING

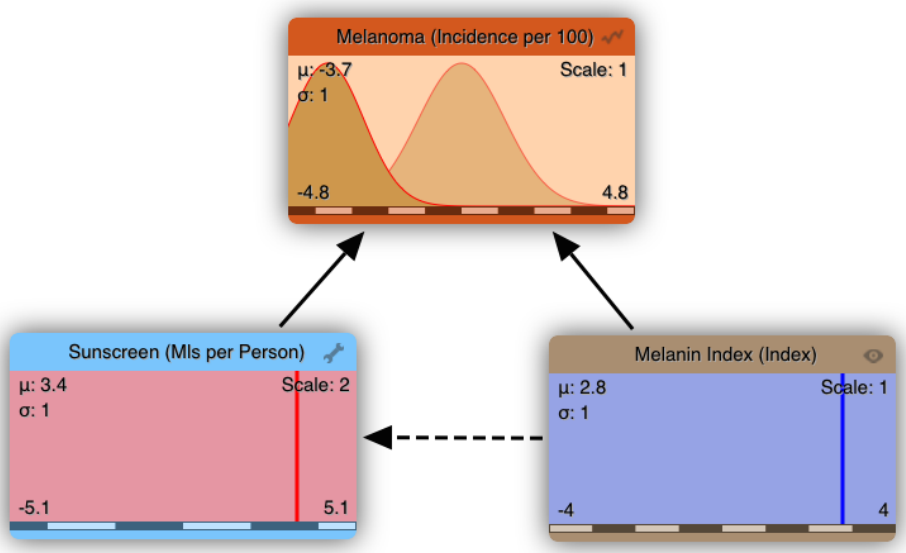
We can approach this in another way: instead of Observing what happens with different distributions of `Sunscreen` and `Melanin Index`, we can try Doing something so that `Sunscreen` is not affected by `Melanin Index`. A real-world analogy would be an experiment or a law that enforced `Sunscreen` use, regardless of skin pigmentation, so that `Melanin Index` is eliminated as a confound. Some people will find the idea of a sunscreen law objectionable and anyone who enacted such a law would have to contend with problems related to compliance and enforcement. But never mind all that! This is just a "what if" scenario: What if

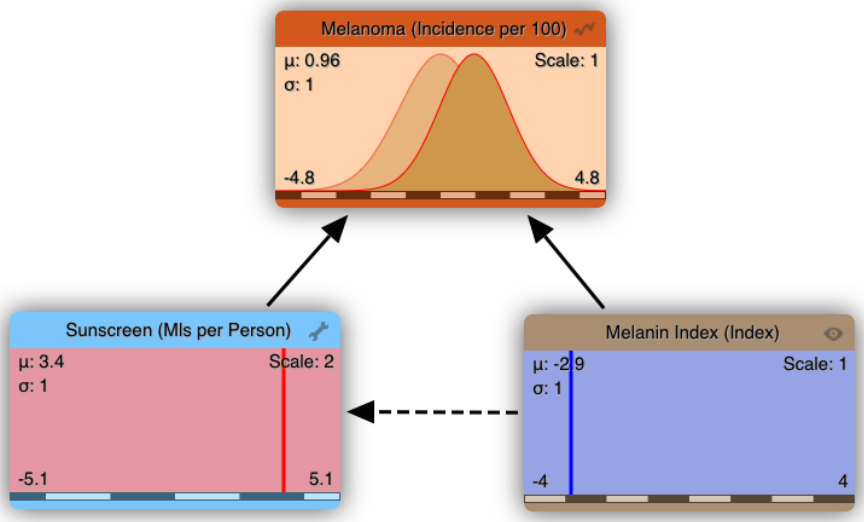
we enforced sunscreen usage? What if sunscreen use had nothing to do with skin pigmentation?

Tap the eye icon a second time to put the Sunscreen variable into Doing mode (indicated by a wrench icon). Doing means we intervene to fix the value of `Sunscreen` in such a way that `Melanin Index` has no effect on it – you will notice that the arrow from `Melanin Index` to `Sunscreen` becomes a dashed arrow. Now drag the observed value line around: As `Sunscreen` use goes down, `Melanoma` increases; as `Sunscreen` use increases, `Melanoma` rates go down. `Melanin Index` is unaffected because we are manipulating `Sunscreen` use directly; there is no relationship between `Sunscreen` use and `Melanin Index` or skin pigmentation.



If you perform these manipulations of `Sunscreen` while Observing different values for `Melanin Index`, you will notice that, while increased Sunscreen use always lowers melanoma rates, the mean for `Melanoma` is higher for people with lighter skin and lower for people with darker skin. `Melanin Index`, in this scenario, has no effect on `Sunscreen` but it is still a factor when it comes to the observed incidence of `Melanoma`: The lighter the skin, the greater the susceptibility to `Melanoma`. However, `Sunscreen` use always reduces the incidence of `Melanoma`.





Again, we can conclude that `Melanin Index` is a confound and that `Sunscreen` use reduces the incidence of Melanoma. But this time we were able to reach that conclusion much more quickly and directly.

How would we calculate $P(M|S)$ in this case? The calculation above (where we are Observing) is not correct here. In this scenario, where we *manipulate* `Sunscreen` in such a way that it does not depend on skin pigmentation, it would an error to multiply $P(M|DS)$ by $P(D|S)$. There *is* a relationship between `Melanin Index` (or, in binary terms, D and ¬D) and `Melanoma` rate but `Sunscreen` use is independent of skin pigmentation. So in the Doing case, $P(M|DS)$ should be multiplied by the proportion of persons who are dark-skinned, and $P(M|\neg DS)$ by the proportion that are light-skinned, which gives us this:

$$P(M|S) = P(M|DS)P(D) + P(M|\neg DS)P(\neg D)$$

In graphical terms, the difference between the Observing calculation and the Doing calculation is the presence (Observing) or absence (Doing) of an arrow between `Melanin Index` and `Sunscreen`.

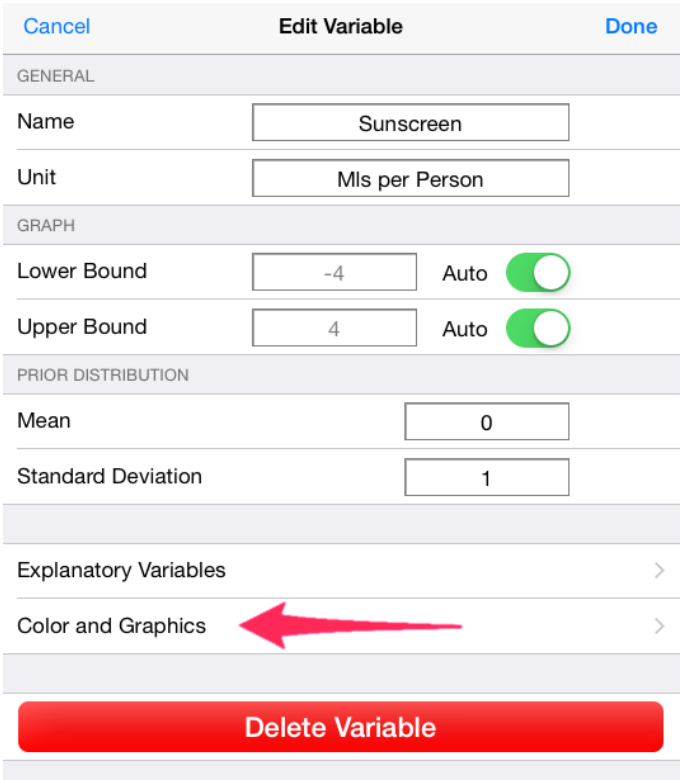
MENU ITEMS

NAVIGATION

While in any menu screen, pressing the Done button located in the top right of the menu bar on will apply any changes made since opening the menu and return you to the network. The button on the upper left of the menu bar will take you back to the previous screen. If the button in the upper left reads Cancel, the menu will be closed without applying the changes made.

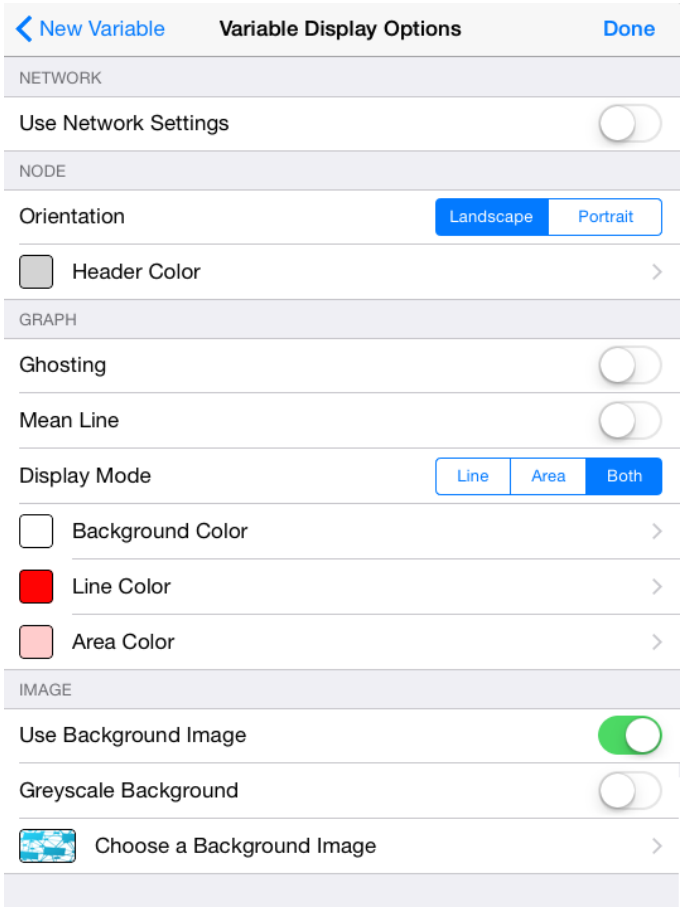
COLOR AND GRAPHICS

To change the display options for a variable, tap and hold the variable, then tap Color and Graphics.



VARIABLE DISPLAY OPTIONS

Turn Use Network Settings off to over-ride the network settings for this variable. (For network settings, see NETWORK DISPLAY OPTIONS below.)



If you over-ride the network settings, you will have access to a number of options that affect the look of each variable. Each option is explained further below:

Orientation – Each node can have either a portrait or landscape aspect ratio. Nodes in portrait mode show a bar graph rather than the normal distribution curve displayed in nodes with landscape orientation.

Header Color – The color of the top and bottom of the node (the area where the name of the variable is displayed).

Ghosting – When Ghosting is turned on, the prior distribution for a variable appears as a ‘ghost’ behind any posterior distribution.

Mean Line – Show or hide a vertical line positioned at the current mean of a variable. This line is only visible for variables in landscape orientation.

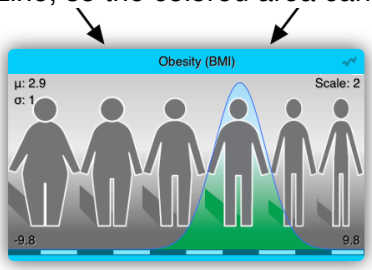
Display Mode – Choose whether the graph is shown as a line, a filled area, or both.

Background Color – The color of the node behind the distribution curve. This color is only shown if a background image is not being used.

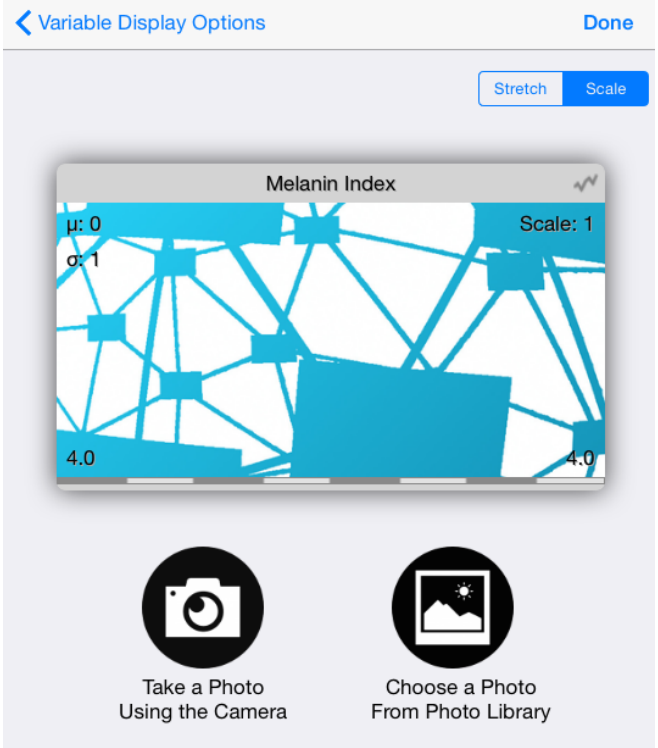
Line and Area Color – The colors of the distribution curve.

Use Background Image – Turn this option on to show an image in the node behind the distribution curve.

Greyscale Background – When this mode is turned on, only the area of the image that is under the distribution curve is shown in color, the rest of the image is shown in black and white (as shown below). This mode is only available for variables in landscape orientation. Note: be sure to set the graph Display Mode to Line, so the colored area can be seen.

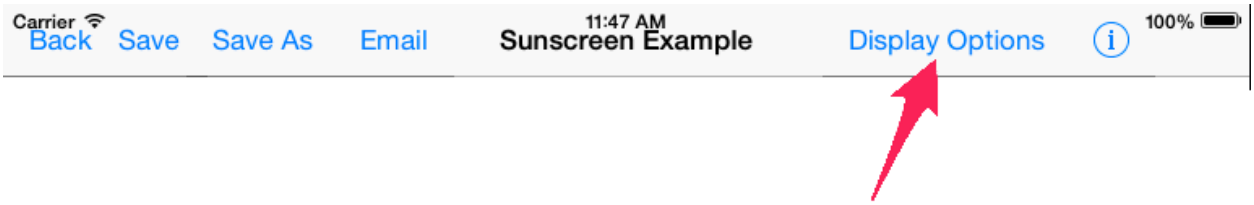


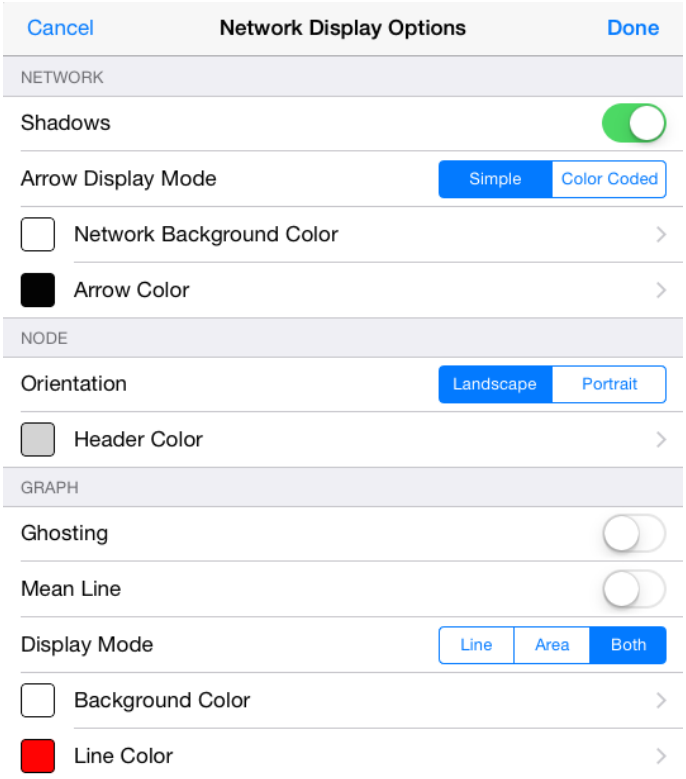
Choose a Background Image – Selecting this option opens a new menu screen that allows you to choose an image to set as the background of a node. A new image can be selected by choosing from your device's photo library, or by capturing an image with the camera. The Stretch and Scale options alter the way the image fits in the node. If Stretch is selected, the entire image will be visible in the node, although, the aspect of the image may be distorted due to the stretching. If the Scale option is selected, the aspect of the image is maintained, but there in no guarantee the entire image will be shown.



NETWORK DISPLAY OPTIONS

To change the display options for a network, tap Display Options in the upper right of the main window.





The display options available for the network are the same as those for individual variables (listed on page 13 and 14), with the addition of:

Shadows – Turn shadows behind nodes on or off.

Arrow Display Mode – Set the Arrow Display Mode to Simple or Color Coded. If you choose Color Coded, the intensity of the color of an arrow will reflect the strength of the relationship between two variables.

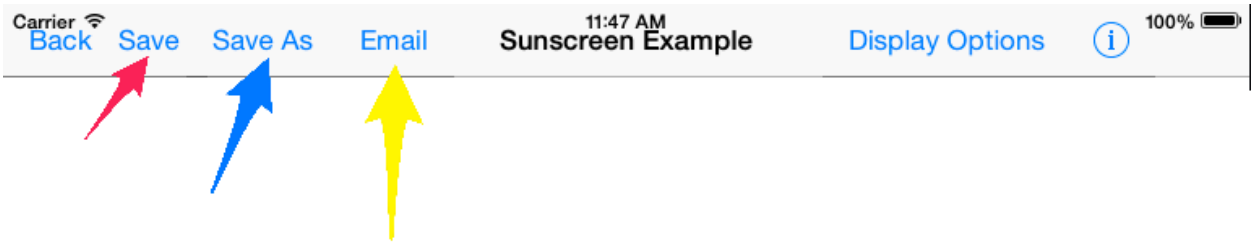
Network Background Color and Arrow Color – Select the color of the background and arrows between variables.

SAVE, SAVE AS, and EMAIL

Tap Save (located in the upper left of the main window) to save any changes that have been made to the network. This option will not be available until the network has a name.

Tap Save As to save a network under a new name. The first time a network is saved, Save As must be selected in order to set a name for the new network.

Tap Email to email a zip file containing the xml file describing a network and any images that are used in the network. The Email button will not be available until the network has been saved under a name using Save As. Note: any unsaved changes will not be sent in the email.



HELP

Tap the ⓘ icon in the upper right corner of the main window to review the gestures described above.